CS230

Vision-Guided Laser Safety System

Aidan J. Fitzpatrick
Department of Electrical Engineering
Stanford University
ajfitz@stanford.edu

Abstract

Advancement in computer vision over the last decade or so has opened the door to new applications exploiting facial recognition and object detection that are being widely researched and more recently implemented in commercial products and systems. The vast pool of existing work in these spaces allows for rather easy expansion into new application spaces by applying transfer learning. In this work, I discuss adapting commonly used deep neural network models for use in a vision-guided laser safety system. This system aims to protect researchers and industry professionals that work with high power lasers. The system performs two functions: 1) ensuring the laser user is properly trained and authorized to use the laser, and 2) detects whether the user is wearing the proper protective eyewear. The real-time outputs of the computer vision modules can be used to determine appropriate control signals to send to a laser system.

1 Introduction

With the growing prevalence and demand for lasers in industrial, commercial, and even medical applications, many researchers and industry professionals are working with lasers and laser systems on a daily basis.

It is no secret that lasers, if used carelessly, can be very dangerous. The American National Standards Institute (ANSI) has studied the impact of optical exposure to the human eye and skin to determine maximum permissible exposure (MPE) limits to prevent eye damage and skin burns in systems that are deployed commercially and in open environments. Often times researchers and industry professionals are using lasers in more controlled settings that significantly exceed these MPE limits. When working with higher power and higher energy laser sources (Class 4), even diffuse optical reflections from within the controlled setup can exceed the ANSI-defined MPE. Therefore, to minimize the risk of adverse biological effects from high levels of optical exposure, proper protective equipment must be worn – particularly protective eyewear due to the extreme vulnerability and ease of damage to the retina.

It is critical to wear protective eyewear every time a Class 4 laser system is operated; however, even advanced researchers can accidentally forget to do so. This unprotected exposure could potentially lead to permanent vision loss. In addition to protective equipment, it is important and required by the Occupational Safety and Health Administration (OSHA) that users of high power or high energy laser systems are properly trained on laser safety practices.

With the dangers of laser exposure in mind, to ensure and enforce accordance with OSHA and ANSI regulations, my project aims to develop a vision-guided laser safety system. This system uses computer vision to detect 1) if the user of the laser is trained (i.e. an authorized user) and 2) if the user is wearing the proper protective eyewear during laser operation. If one or both of these requirements

CS230: Deep Learning, Fall 2020, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

are not satisfied, the system will disable the power supply units of the laser to prohibit use. A diagram of the laser safety system is shown in Fig. 1.

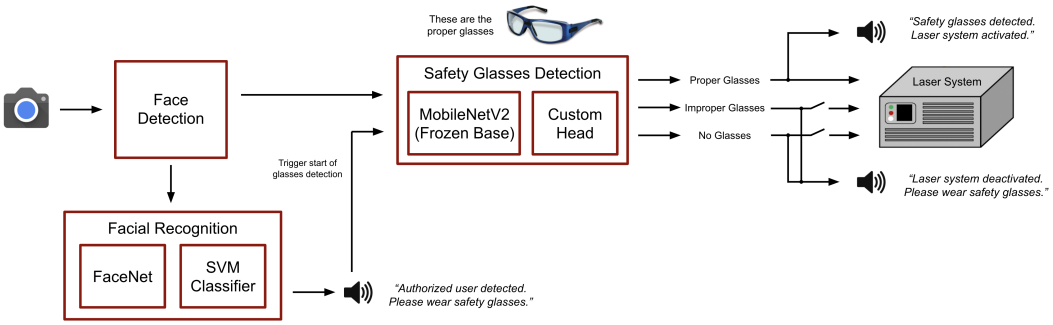

Figure 1: Block diagram of the system concept showing 3 key subsystems.

The chain consists of two computer vision models to perform the individual tasks and one common model for face detection. The first is a pretrained FaceNet (1) model to perform facial recognition to determine if the user is an authorized user; this model takes a video frame as input, computes an embeddings vector from the image, and finally passes these embeddings to a trained classifier model (a linear support vector machine (SVM)) to predict the user's identity. If the user is defined in the system as an authorized user, the system audibly prompts the user to wear safety glasses. At this point, another video frame is read and passed as input to the second model of the system. The second model is a MobileNetV2 (2) model which has been augmented to perform detection of laser safety glasses. Further details on the implementation, results, and future work of this project are outlined in the remainder of this report.

2 Methods

In this section, I describe the chain depicted in Fig. 1 piece-by-piece to articulate the importance and implementation of each subsystem.

2.1 Face Detection

As seen in Fig. 1, the first block in the system chain is to perform face detection on the current video frame. The face detection is performed with a commonly used pretrained Caffe-based face detector model that makes use of the Single Shot MultiBox Detector framework (3). This model was chosen due to its advantages over other face detection techniques: its high frame rate, its robustness to quick movements, its ability to handle occlusions well, and its ability to identify side faces. These are important features for a face detection system particularly in this application since it is desired to operate in real-time in an environment where the user will be moving around the lab performing other tasks.

2.2 Facial Recognition

The next block performs facial recognition on the faces detected. Facial recognition is typically done by training a neural network to calculate an embedding (or a feature vector) for the input image rather than classifying the identity directly. This feature vector can then be passed to a classifier to determine the identity. This two-step approach is due to the fact that it is difficult to collect a large enough dataset of images for each of the identities you would like to recognize. Instead, the neural network can be trained to simply represent the input image uniquely in a high-dimensional feature space. A classifier, such as a support vector machine, can then be trained on a more limited dataset to separate the identities encoded in the feature space.

A common facial recognition deep neural network uses the FaceNet architecture, which is based on the Inception model and makes use of residual connections to significantly accelerate the training process (4). The FaceNet model used in this work, which computes a 512-dimensional feature vector, was pretrained using the VGGFace2 dataset (5). The classifier that follows the FaceNet model, which

in this work is a linear SVM, must only be trained on identities which are desired to be recognized. The training of the classifier is therefore very fast and efficient and can be updated to add identities fairly easily whereas retraining the entire FaceNet model requires a lot of computational resources.

To train the classifier, I use a dataset of 10-25 images per authorized user; this is sufficient since the feature vectors for these classes are far from each other in the high-dimensional vector space. To show an example, Fig 2 illustrates a 2-component principal component analysis on the embeddings generated by FaceNet for 3 different identities. It is clear in this figure that the embeddings are highly clustered for each user and easily separable with linear decision boundaries.

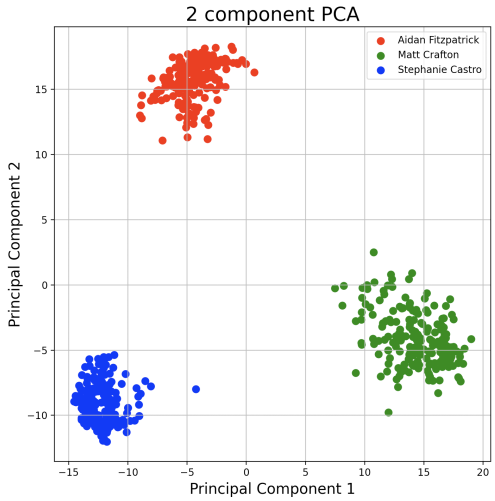

Figure 2: 2-component principal component analysis on the embeddings generated by FaceNet for 3 different users.

If the embedding generated by FaceNet is not close in the Euclidean space to any of embeddings in the classifier training set, the identity is classified as an unauthorized user of the laser and the system does not advance to the safety glasses detection model.

2.3 Safety Glasses Detection

Once an authorized user is detected, a current video frame is captured and passed through the face detector model to acquire all faces in the frame which may or may not be wearing the proper protective eyewear. The safety glasses detection model is responsible for making this distinction.

Detection of the laser safety glasses can be classified as an object detection problem. In a literature survey for object detection, one can find many architectures that demonstrate high performance (6). In this work, I choose to use the computationally efficient MobileNetV2 architecture (2) with hopes to eventually migrate this system to an edge device such as a Raspberry Pi module.

The computational efficiency of the MobileNetV2 architecture comes by virtue of replacing standard convolutional layers with depthwise separable convolutions (2; 7). A depthwise separable convolution replaces a standard convolution with a factorized version by first performing a depthwise convolution (or applying a single filter per channel) and then a 1×1 convolution (2). The depthwise separable convolutional layers as implemented by the creators of MobileNetV2 reduces the computational cost by about 9 times while only sacrificing a small reduction in accuracy. Further details of the MobileNetV2 architecture can be found in (2).

To adapt the MobileNetV2 architecture for use in my laser safety glasses detection application, I apply transfer learning on a model that has been pretrained on the ImageNet database (8). By freezing the base of the model and redesigning and retraining the model's head, the MobileNetV2 architecture can be employed specifically for the desired task. The output of the glasses detector model is one of three classes: 1) proper safety glasses, 2) improper safety glasses, or 3) no safety glasses. Note that a prediction is only made if a face was detected by the face detector; therefore, it is not necessary to have a "background" class. With three output classes, the final layer of the model has three hidden

units and makes use of a softmax activation function such that the output provides a probability for each class.

For training and testing, I captured a total of 2575 images with the following distribution: 866 proper safety glasses, 818 improper glasses, and 891 no safety glasses. The head to be retrained consists of an average pooling layer with pool size of 7×7 to reduce the dimensionality before flattening, a fully connected layer implemented with a 50% dropout probability for regularization, and finally the softmax output layer. For training, I use a batch size of 32, Adam optimization with a categorical cross-entropy loss function, and an initial learning rate of 1e-4 with learning rate decay over 20 epochs. 20% of the data was withheld for cross-validation demonstrating greater than 99.8% classification accuracy. This was sufficiently high performance for the application and thus the hyperparameters and optimization techniques were not studied in-depth.

Finally, it is important to note that this system runs continuously and in real-time such that if a user removes their safety glasses during laser operation, the laser is immediately deactivated.

3 System Demonstration

In this section, I provide system outputs to demonstrate its functionality. Firstly, in Fig. 3 the outputs to the facial recognition model are shown for an authorized user versus an unauthorized user.

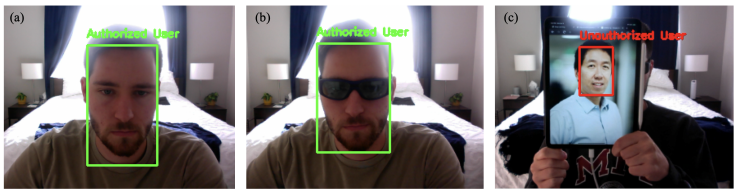

Figure 3: Video frame outputs for (a) an authorized user, (b) an authorized user wearing safety glasses, and (c) an unauthorized user.

Following detection of an authorized user, there exists an audible output which states: *"Authorized user detected. Please wear safety glasses."* It is shown in the figure that an authorized user can be detected regardless of whether the safety glasses are being worn. The facial recognition model nor the subsequent SVM were trained on images of authorized users wearing safety glasses. To investigate how this is possible, I again examine a 2-component principal component analysis on the embeddings generated by FaceNet but this time including images with and without safety glasses.

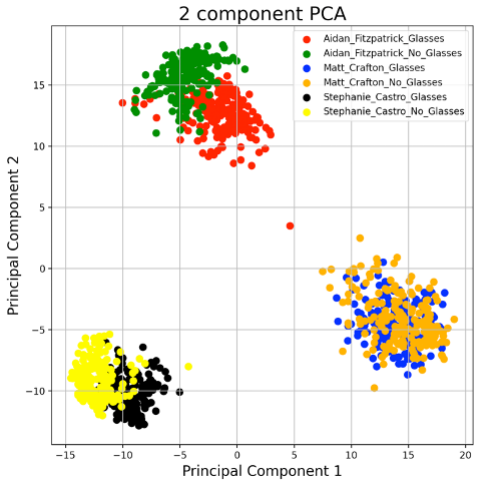

Figure 4: 2-component principal component analysis on the embeddings generated by FaceNet for 3 different users wearing and without wearing safety glasses.

As shown in Fig. 4, the embeddings overlap significantly for a user wearing safety glasses and not wearing safety glasses. This means that the trained FaceNet model is not sensitive to slight obstructions to the eyes and produces similar embeddings regardless of these two classes. An optimistic view is that this is favorable for my system as it is highly generalizable to different types of safety glasses and can be trained simply on images of authorized users without safety glasses being worn. However, the fact that these are not separable is actually what requires us to have an additional model for safety glasses detection.

Next, in Fig. 5 the outputs to the glasses detector model are shown.

Figure 5: Video frame outputs for (a) safety glasses detected, (b) improper safety glasses detected, (c) no safety glasses detected, and (d) one user with safety glasses and one user without safety glasses.

Similarly to above, there exists an audible output for each of the above cases. First, in Fig. 5(a) the output states: *"Safety glasses detected. Laser system activated."* For the cases shown in Fig. 5(b)-(d), the output states: *"Laser system deactivated. Please wear safety glasses."* In Fig. 5(b), the system informs the user that improper safety glasses are being worn by labeling the bounding box of the face detection. In Fig. 5(c), no safety glasses are being worn as shown by the bounding box label. Finally, for the case in Fig. 5(d), the output also indicates to deactivate the laser system. This is to ensure that all users in proximity to the laser are wearing the proper eyewear.

Since this system is run in real-time, the audible outputs only occur when there is a change of laser state. To prevent constant laser switching and perhaps experimental interruption for false detections, the system requires 5 subsequent frames of the same label before changing state. The resulting latency here is practically negligible due to the high frame rate of the face detection.

4 Conclusion

For this project, I re-imagine laser safety by making use of the recent and rapid advancements in the field of computer vision. The designed laser safety system performs facial recognition to confirm that the user has completed the appropriate laser safety training. In addition, the system detects the use of proper protective eyewear. In the event that one or both of these conditions are not satisfied, the laser safety system disables the laser and prohibits its use.

This system is fully operational in an independent capacity as demonstrated in Section 3 and will soon be integrated with a true laser system to enact the appropriate controls. Additional future work will include porting the models, possibly in a lighter weight implementation, to an edge device such as a Raspberry Pi so that the system can operate continuously in a stand-alone fashion and have no dependency on a personal computer. Finally, it is desired to augment the facial recognition with liveness detection to prevent authorized user spoofing.

References

[1] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.

[2] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[3] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[4] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," *arXiv preprint arXiv:1602.07261*, 2016.

[5] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 67–74.

[6] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[7] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.

[8] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.